Zoosyst. Evol. 100 (4) 2024, 1297–1314 | DOI 10.3897/zse.100.127201

PENSOFT

für Natur MUSEUM FÜR NATURKUNDE BERLIN

Molecular systematics of *Perinereis* and an investigation of the status and relationships of the cultured species *Perinereis wilsoni* Glasby & Hsieh, 2006 (Annelida, Nereididae)

Deyuan Yang[1,2], Sheng Zeng[2], Zhi Wang[3], Yanjie Zhang[4], Dazuo Yang[5], Christopher J. Glasby[6,7], Jiang-Shiou Hwang[1,8], Lizhe Cai[2]

1 *Institute of Marine Biology, National Taiwan Ocean University, Keelung, Taiwan*
2 *College of the Environment and Ecology, Xiamen University, Xiamen 361102, China*
3 *State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China*
4 *School of Life and Health Sciences, Hainan University, Haikou, China*
5 *Key Laboratory of Marine Bio-resource Restoration and Habitat Reparation in Liaoning Province, Dalian Ocean University, Dalian 116023, China*
6 *Natural Sciences, Museum & Art Gallery Northern Territory, PO Box 4646, Darwin, NT 0801, Australia*
7 *Australian Museum Research Institute, 1 William Street, NSW 2010, Sydney, Australia*
8 *Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan*

https://zoobank.org/95987DFE-0752-4899-B07D-BC6C93DC8C60

Corresponding authors: Lizhe Cai (cailizhe@xmu.edu.cn); Jiang-Shiou Hwang (jshwang@mail.ntou.edu.tw); Christopher J. Glasby (chris.glasby@nt.gov.au)

Academic editor: Greg Rouse ♦ **Received** 9 May 2024 ♦ **Accepted** 16 August 2024 ♦ **Published** 13 September 2024

Abstract

In this study, we conducted morphological and molecular analyses of *Perinereis wilsoni*, a species being considered for aquaculture in China. We found this species difficult to identify because of its close morphological similarity to the sympatric *P. mictodonta* and thus sought genetic markers to more easily distinguish it and to investigate its phylogenetic relationship to *P. mictodonta* and other nereidids. For the first time, we sequenced, assembled, and annotated the complete mitochondrial genome, nuclear ribosomal sequences (*18S-ITS1-5.8S-ITS2-28S*), and four nuclear histone genes (*H3-H2A-H2B-H4*) of *P. wilsoni*. Comprehensive bioinformatics methods were employed to assemble the genome-skimming data of *P. wilsoni* to ensure assembly quality. Phylogenetic analyses based on five datasets of the available mitochondrial genomes (32 taxa in Nereididae, including 8 taxa in *Perineris*), using maximum likelihood and Bayesian analyses, provide support for the monophyly of the genus *Perinereis*. In contrast, the *P. nuntia* species group, a subgroup within *Perinereis*, is nonmonophyletic. *Perinereis wilsoni* has a closer phylogenetic relationship with *P. vancaurica* and *P. nuntia*. Our study serves as a baseline for future work on the cultivation, reproductive biology, and phylogeny of *P. wilsoni*.

Key Words

Bioinformatic analyses, genome skimming, mitogenomes, *Perinereis*

Introduction

Recently, a local *Perinereis* species intended for large-scale aquaculture in China by the Key Laboratory of Marine Bio-resource Restoration and Habitat Reparation in Liaoning Province, Dalian Ocean University, was identified as *P. wilsoni* Glasby & Hsieh, 2006. The genus *Perinereis* Kinberg, 1865, belongs to the Nereididae. This family comprises over 700 described species and 45 genera (Wilson et al. 2023). Many nereidid species, particularly *Perinereis*, are of significant commercial and ecological importance for fishing bait, aquaculture feed, and wastewater treatment (Palmer 2010; Arias et al. 2013). In China, *Perinereis aibuhitensis* (Grube, 1878) is farmed

Copyright *Yang, D. et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and exported worldwide as fish bait, and *Tylorrhynchus heterochetus* (Quatrefages, 1866) is consumed as food by people in Guangdong Province and other Southeast Asian countries (Glasby and Hsieh 2006; Wilson et al. 2023).

Perinereis represents the second-largest genus in Nereididae, with approximately 100 species (Mahcene et al. 2023; Prajapat et al. 2023; Wilson et al. 2023; Teixeira et al. 2024). Bakken and Wilson (2005) conducted a phylogenetic analysis based on morphological evidence, suggesting that *Perinereis* may be polyphyletic. This was supported by subsequent molecular studies (Glasby et al. 2013; Alves et al. 2020; Elgetany et al. 2022). For practical identification purposes, Hutchings et al. (1991) first divided *Perinereis* into three species groups based on the number of pharyngeal Area VI paragnaths and further divided each group based on parapodial characters. The *Perinereis nuntia* species group was characterized by a distinctive arc of bar-shaped paragnaths on area VI (usually numbering 6–14 on each side) (Hutchings et al. 1991; Wilson and Glasby 1993; Glasby and Hsieh 2006). This group has been reviewed by Wilson and Glasby (1993), Glasby and Hsieh (2006), and Villalobos-Guerrero (2019). Currently, it comprises approximately 20 recognized species (Villalobos-Guerrero 2019; Wilson et al. 2023), with the type localities of about 13 species in the Indo-West Pacific. These studies have significantly improved our understanding of this group. However, relying solely on a morphology-based approach may not be sufficient to resolve the taxonomic status of these species because some key taxonomic characters used to distinguish species often overlap, as they show ontogenetic and intraspecific variation (Tosuji et al. 2019; Tosuji et al. 2023). These morphology-defined species need to be re-evaluated using integrated taxonomic methods by taxonomic specialists. Recently, studies based on integrative taxonomy have revealed more new species in Nereididae (see Glasby et al. 2013; Teixeira et al. 2022a; Teixeira et al. 2022b).

Perinereis wilsoni, a member of the *P. nuntia* species group, is morphologically very similar to *Perinereis mictodonta* (Marenzeller, 1879), with no distinct morphological differences, although slight but statistically significant morphometric differences were found with respect to paragnath numbers and the relative length of the dorsal cirri (Glasby and Hsieh 2006). For this reason, these two species were primarily established based on the differences in *ITS* (internal transcribed spacer) genes (Chen et al. 2002; Glasby and Hsieh 2006). Therefore, identifying these two species based solely on morphology presents a significant challenge. To date, only two studies have contributed molecular data: Chen et al. (2002) provided *ITS* genes. Tosuji et al. (2019) provided sequences for both the partial mitochondrial *16S* ribosomal RNA (*16S*) and *ITS* genes. Although Park and Kim (2007) researched *P. wilsoni* using the partial mitochondrial cytochrome oxidase I (*COX1*) gene, these sequences were not made publicly available. Moreover, there are some *COX1* sequences labeled *P. wilsoni* in the NCBI GenBank database, but these sequences remain pending verification.

Given the considerable economic importance of *Perinereis* species, accurate species identification is crucial because correct scientific names can facilitate the linking of subsequent physiological or reproductive studies, thus ensuring the reproducibility of these research efforts (Glasby and Hsieh 2006; Hutchings and Lavesque 2020). Therefore, we have focused on several key questions regarding taxonomy: 1) Is the identification of *P. wilsoni* from Liaoning Province accurate? 2) Which molecular markers are suitable for identifying this species? 3) What is the phylogenetic relationship of *P. wilsoni* to other nereidids?

In this study, we used low-coverage whole genome sequencing, also known as genome skimming (Straub et al. 2012), which has been widely used in polychaete phylogeny studies because it is cost-effective and does not require fresh tissue (Richter et al. 2015; Coissac et al. 2016; Alves et al. 2020; Hektoen et al. 2024). Utilizing this strategy, we provided more molecular information, including high-copy regions: mitochondrial genome (mitogenome), nuclear ribosomal sequences (*18S* rRNA-*ITS1-5.8S* rRNA-*ITS2-28S* rRNA), and nuclear histone genes (*H3-H2A-H2B-H4*).

Materials and methods

Sample collection, identification, and sequencing

The specimens of *P. wilsoni* were sampled from Dalian, Liaoning Province, China (38.8732°N, 121.6767°E) and identified by Deyuan Yang. All specimens were fixed directly in 95% ethanol. Two specimens with a fully-everted pharynx were used for morphological and molecular studies. They were deposited at Xiamen University (XMU) under voucher numbers 23007-1 and 23007-2, respectively. Specimens were identified based on the key in Glasby and Hsieh (2006). Methods used in the morphological study were followed by Yang et al. (2022). In summary, whole worms were photographed with an Olympus E-M1 Mark II camera with a 60 mm macro lens, detailed structures with a Zeiss Discovery V20 stereomicroscope, and a Nikon 80i compound microscope. Image stacks were obtained using Helicon Focus 7 (https://www.heliconsoft.com/heliconsoft-products/helicon-focus/) and post-processed using Adobe Photoshop. The terminology of nereidid followed Wilson et al. (2023) generally. To better describe areas II, III, and VI of paragnaths, we introduced the notations 'L:R' and 'L:M:R'. 'L:R' is for areas II and VI, and 'L:M:R' to describe the paragnath patch in area III, where 'L' represents the number of paragnaths on the left, 'M' in the middle, and 'R' on the right. For example, for area II, '2:3' signifies 2 paragnaths on the left and 3 on the right; for area III, '3:5,2R:7' signifies 3 paragnaths on the left, a central patch with 5 paragnaths in 2 rows, and 7 paragnaths on the right.

Before DNA extraction, each individual was cleaned with 95% ethanol. To avoid gut contamination, two to seven parapodia were clipped from the specimens. Whole genomic DNA was extracted using the TIANamp Genomic DNA Kit (TIANGEN, Beijing, China). Genome skimming was conducted on the Illumina NovaSeq X Plus with a PE150 strategy. DNA extraction and sequencing were performed at Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). Our initial sequencing effort was to obtain 5 Gb of raw data for each sample. Voucher number 23007-2 was increased to 25 Gb to explore more universal single-copy ortholog genes.

Assembly and annotation of mitochondrial and nuclear sequences

Raw paired-end reads were removed from sequence adapters, and low-quality regions were trimmed using Fastp v.0.23.4 (Chen et al. 2018). The quality of all cleaned reads was checked with FastQC v.0.12.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and subsequently, FastQC results were summarized using MultiQC v.1.8 (Ewels et al. 2016) (Suppl. material 1).

To ensure reliable assembly for the present data, various assemblers were used: 1) SPAdes v.3.15.5 (Bankevich et al. 2012), Ray v.2.3.1 (Boisvert et al. 2010), Megahit v.1.2.9 (Li et al. 2015), and IDBA-UD v.1.1.3 (Peng et al. 2012); 2) GetOrganelle v.1.7.6.1 (Jin et al. 2020); 3) MITObim v.1.9 (Hahn et al. 2013) and NovoPlasty v.4.3.1 (Dierckxsens et al. 2017), with *16S* (GenBank accession numbers: LC482188) as the seed; 4) our new assembly pipeline, FastMitoAssembler (original version at https://github.com/suqingdong/FastMitoAssembler). The pipeline implemented in FastMitoAssembler involved the following steps: First, MEANGS v.1.2.1 (Song et al. 2022) was used to obtain and assemble mitochondrial genes, which then served as a seed input for NovoPlasty. Second, the results of NovoPlasty served as a seed input for GetOrganelle. Third, MitoZ v3.6 (Meng et al. 2019) was used for annotation. This workflow was managed using Snakemake (Köster and Rahmann 2012), which can handle large-scale samples with minimal input–only raw data.

For *de novo* assembly results (SPAdes, Megahit, Ray, and IDBA-UD), we used MitoFinder v.1.4 (Allio et al. 2020) to extract mitochondrial contigs (mt contigs) and annotations, using the *Platynereis dumerilii* mitogenome (GenBank: NC 000931) as a reference. Unfortunately, no mt contigs were found from the Ray assembler. If the circularization of a mt contig was not automatically completed by MitoFinder, the circules.py script (https://github.com/chrishah/MITObim/tree/master/misc_scripts) was used to assess the circularization of the contig. After obtaining mitogenomes from various assemblers, Quast v.5.2.0 (Gurevich et al. 2013) was employed for assembly quality evaluation (Suppl. material 2). A BLASTn query against GenBank in the NCBI (NCBI BLAST Analysis, https://BLAST.ncbi.nlm.nih.gov/BLAST.cgi) was also performed to check for contamination in the mitogenomes (Suppl. material 3: table S1). The GC content and read (per-base) coverage depth of mitogenome sequences were calculated using the visualize subcommand in MitoZ v.3.6 with default parameters.

The mitogenome was annotated using MITOS2 (Donath et al. 2019), MitoFinder v.1.4.1, and MitoZ v.3.6. All annotation files (in .gb or .gbf formats) were reordered with COX1 as the start gene using PhyloSuite v.1.2.3 (Zhang et al. 2020). The reordered GenBank format (gb) file was converted to a fasta file and uploaded to GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) to combine the results of ARWEN v.1.2.3 and tRNAscan-SE v.2.0.7 into a single gb file. This approach allowed us to compare tRNA annotations generated from three common methods (ARWEN, tRNAscan-SE, and MiTFi). All annotations were then loaded into Geneious Prime v.2022.2.2 for manual curation. First, we utilized the MAFFT plugin in Geneious to align these gb files. Then, we checked the coding region annotations with the assistance of the ORFs function based on the invertebrate mitochondrial genetic code table 5. Second, tRNAs were evaluated using the following criteria: tRNAscan-SE was given precedence over ARWEN and MiTFi, while MiTFi had precedence over ARWEN. Thus, if tRNAscan-SE identified a tRNA, its result was chosen. In cases where both MiTFi and ARWEN identified a tRNA, the result from MiTFi was selected. Subsequently, we imported other gb files of *Perinereis* into Geneious to compare each gene and manually edit them.

The boundaries of each gene were determined by the following rules: 1) Protein-coding genes can overlap with each other but cannot overlap with tRNA genes. In Nereididae, only the *ND4* and *ND4L* coding genes are allowed to overlap. 2) The boundaries of mitochondrial rRNA genes (*12S* and *16S*) were defined by flanking genes. 3) The large non-coding region (or putative control region) was determined by the boundaries of neighboring genes and a low GC region.

The CGView Comparison Tool (CCT) (https://github.com/paulstothard/cgview_comparison_tool) (Grant et al. 2012) was employed to compare all accessible mitogenomes of Nereididae in NCBI (as of 15 January 2024) to evaluate the quality of annotations further.

The high-copy nuclear markers (*18S* to *28S* genes and histone genes) were assembled using GetOrganelle v.1.7.6.1. Considering the lack of effective software or pipelines to annotate these genes, we manually annotated nuclear rRNA genes in Geneious using primer pairs to define gene boundaries: 18*SA* and *18SB* (Medlin et al. 1988) for *18S*, *F63.2* and *R3264.2* (Struck et al. 2006) for *28S*, and *ITS18SFPOLY* and *ITS28SRPOLY* (Nygren et al. 2009) for the *ITS1-5.8S-ITS2* region. After completing one sequence annotation, we utilized the "Transfer annotations" function within Geneious Prime v.2022.2.2 to annotate other sequences. The histone genes were annotated using the same function, with GenBank accession

number X58895 (*Urechis caupo*) as a reference at an 85% similarity threshold. If the number of gene annotations was incomplete in one sequence, we imported the corresponding cleaned sequencing reads into Geneious and reassembled the sequence employing the "map to reference" function in Geneious Prime v.2022.2.2 (mapper: Geneious, fine-tuning: iterate up to 25 times) and then re-annotated it. Once annotations were completed, each gene was extracted for NCBI BLAST analysis. The read (per-base) coverage depth of nuclear genes was calculated using a custom script, which used BWA (Li 2013) and SAMtools (Li et al. 2009).

BUSCO v.5.6 was employed to generate universal single-copy orthologs (USCOs) (Simão et al. 2015) using only the assembly results (contigs) from SPAdes with the following parameters: -l metazoan_odb10, -m genome, --augustus, -c 40. A total of 618 single-copy genes were searched.

Phylogenetic analyses

The phylogenetic analysis utilized two types of datasets: complete or near-complete mitogenomes of Nereididae and DNA barcoding sequences (*COX1*, *16S*, and *ITS*) of *Perinereis*. All publicly available Nereididae mitogenomes from GenBank (as of 15 January 2024), along with two outgroup species (*Craseoschema thyasiricola* NC 060815 and *Leocrates chinensis* NC 066969, belonging to Chrysopetalidae and Hesionidae, respectively), were included in the phylogenetic analysis. The outgroup taxa were selected based on previous hypotheses that Chrysopetalidae and Hesionidae are sister groups to Nereididae (Glasby 1993; Pleijel and Dahlgren 1998; Tilic et al. 2022). To ensure consistent annotation criteria and to correct potential errors in previous annotations of mitogenomes, we re-annotated all available Nereididae mitogenomes using the methods described above. Ultimately, *Platynereis massiliensis* (NC 051996) and *Alitta succinea* (NC 051993) from the Nereididae were excluded from this study (Suppl. material 3: table S2).

To identify potential taxonomically mislabeled sequences in the mitogenomes dataset, we extracted the *COX1* and *16S* genes and conducted NCBI BLAST analysis. The results were downloaded as CSV files. For each sequence, the top 10 matching sequences were sorted based on identity using WEKits v.1.0 (https://github.com/GP-sir/wekits/releases). Next, we retrieved the source information (subjectAcc) for these matched sequences from NCBI. The species identification of each sequence was then manually verified in Microsoft Excel. Specifically, we focused on matched sequences with identity over 97% and checked whether they were supported by reliable taxonomic references (Suppl. material 3: table S3, S4).

All manually curated mitogenomes were imported into PhyloSuite v.1.2.3. The extracted protein-coding genes (PCGs) and two rRNAs (2R) from these sequences were aligned using MAFFT in normal mode. MACSE v.2 (Ranwez et al. 2018) was used to improve the multiple sequence alignment of 13 PCGs in the "refinement strategy." After completing the alignment tasks, each alignment was manually checked in Geneious. Ambiguously aligned regions were removed using Gblocks with the default settings or trimAl v.1.2 (Capella-Gutiérrez et al. 2009) with the automated1 setting (Suppl. material 3: table S5). Five mitogenomic datasets were used in the phylogenetic analyses: (i) 13PCGs: PCGs with all three codon positions; (ii) 13PCGs12: PCGs without 3rd codon positions; (iii) 13PCGs_2R: PCGs dataset plus two rRNA genes; (iv) 13PCG12_2R: PCG12 dataset plus two rRNA genes; and (v) 13PCGsAA: amino acid sequences translated from 13PCGs. The heterogeneity of each dataset was assessed using AliGROOVE v.1.08 (https://github.com/Patrick-Kueck/AliGROOVE) with the default settings.

ModelFinder v.2.2.0 was used to select the best substitution models for each partition of the maximum likelihood (ML) and Bayesian inference (BI) analyses (Suppl. material 3: table S6). ML analysis was carried out in IQ-TREE v.2.2.2 under the edge-linked partition model with 20,000 ultrafast bootstraps. BI analysis was performed using MrBayes v.3.2.7a (Ronquist et al. 2012), with two parallel runs for at least 2,000,000 generations, ensuring the average standard deviation of split frequencies was below 0.01 (Ronquist et al. 2012).

Considering the mitogenomes of *Perinereis* are limited, shorter DNA sequences (DNA barcodes) were also used to explore the phylogenetic position of *Perinereis wilsoni*. All available *Perinereis* genes from NCBI (https://www.ncbi.nlm.nih.gov/nuccore/?term=Perinereis) were downloaded. The *COX1*, *16S*, *18S*, *28S*, *H3*, and *ITS* genes were further extracted because these molecular markers are widely employed in Annelida. To organize the data more efficiently, we used a custom script (Suppl. material 4) to categorize these data based on voucher numbers (treating isolates, clones, strains, and haplotypes as synonymous with voucher numbers). Ultimately, the *COX1*, *16S*, and *ITS* genes were selected for phylogenetic analysis due to their higher representation among taxa, and new sequences from this study were added to explore the phylogenetic position of *Perinereis wilsoni* (Suppl. material 3: table S7). The trees constructed with the *COX1*, *16S*, and *ITS* genes used the same methods as the mitogenome data.

Multiple alignment summary statistics were generated using the alignment summary function in BioKIT (Steenwyk et al. 2022), including the number of taxa (sequences), alignment lengths, number of constant sites, number of parsimony informative sites and variable sites, and the frequency of all character states.

Pairwise tree structure comparison was conducted using the *all.equal.phylo* function in the ape v.5.7.1 package (Paradis and Schliep 2019), and topological differences were assessed using TreeSpace (Jombart et al. 2017), both implemented in R v.4.3.1 (R Core Team 2023). Finally, iTOL v.6 (Letunic and Bork 2021) was used to visualize the trees.

Sequence analyses

Strand asymmetries were calculated using the following formulae (Perna and Kocher 1995): AT-skew = (A - T) / (A + T); GC-skew = (G - C) / (G + C). Codon usage and relative synonymous codon usage (RSCU) of 13 PCGs were computed in PhyloSuite and visualized using the 'ggplot2' package (Wickham 2016) in R v.4.1.3 (R Core Team 2023). DnaSP v.6.0 (Rozas et al. 2017) was used to calculate the non-synonymous (Ka)/synonymous (Ks) substitution rates among the 13 PCGs of Nereididae and the nucleotide diversity (Pi) with a sliding window of 100 bp and a step size of 25 bp. A Ka/Ks ratio < 1 indicates purifying selection, while a ratio > 1 indicates positive selection.

To determine which molecular marker is reliable for accurate species delimitation, all labeled *P. wilsoni* and *P. mictodonta* *COX1*, *16S*, and *ITS* genes in NCBI were used. We used the following criteria: 1) whether the presence of a barcode gap: the minimum interspecific genetic distance is greater than the maximum intraspecific genetic distance; 2) whether each species is recovered as monophyletic; and 3) whether there is a small overlap between intra- and interspecific distances and a large barcode gap, as recommended by dos Santos Vieira et al. (2020). Genetic distances (p-distance and Kimura 2-parameter) over sequence pairs were calculated in MEGA X (Kumar et al. 2018). The heat maps were generated in TBtools v.2.080 (Chen et al. 2020).

Results

Morphological analyses

Perinereis wilsoni Glasby & Hsieh, 2006

Fig. 1A–K

Material examined. • 23007-1 and 23007-2, both collected from Dalian, Liaoning, China (38.87315°N, 121.676671°E), 08 August 2023, preserved in 95% ethanol.

Description. Description based on 23007-1 and 23007-2. 23007-1 complete, 93 chaetigers, 7.0 cm in length, 2.9 mm wide at chaetiger 10 (excluding parapodia); and 4.0 mm wide at chaetiger 10 (including parapodia) (Fig. 1A). 23007-2 complete, 6.5 cm in length, 98 chaetigers, 2.3 mm wide at chaetiger 10 (excluding parapodia), 3.7 mm wide at chaetiger 10 (including parapodia).

Prostomium and anterior dorsum with dark brown pigmentation. Prostomium anterior margin entire, pear-shaped, wider than long, with shallow longitudinal groove in central area. Antennae conical, about one-third length relative to prostomium length. Palps longer than prostomium, biarticulate, with palpophores and palpostyles (Fig. 1A, C, J). Palpophores barrel-shaped, slightly wider basally, palpostyles spherical (or globular). Eyes black (Fig. 1A).

Pharynx fully everted (Fig. 1A, C–F, J–K). Jaws brown, with 7 teeth (based on 23007-1) (Fig. 1B). Tentacular cirri extend back 3–12 setigers; posterodorsal one extending to chaetiger 8–12.

Paragnath counts (Fig. 1C–E, J–K): area I with 1 conical paragnath; area II with 9 or 11 conical paragnaths on left, 11 or 14 conical paragnaths on right (23007-1, II = 9:11; 23007-2, II = 11:14); area III with 14–17 conical paragnaths, central patch with 9–12 in 2 rows, 2–3 laterally on either side (23007-1, III = 3:9,2R:2; 23007-2, III = 2:12,2R:3); area IV with 18–22 conical paragnaths on each side (23007-1, IV = 18:25; 23007-2, III = 22:22), bars absent; area V with 1 conical paragnath; area VI with 3–5 shield-shaped bars on each side (23007-1, VI = 5:3; 23007-2, VI = 5:5); area VII–VIII with 19 conical paragnths in 2 rows.

Notopodia with 2 lobes, prechaetal lobe absent. Dorsal cirri longer than notopodial ligule, about 1.5 times length of the notopodial ligule throughout. Notopodial ligule similar to median ligule throughout. Neuropodial postchaetal lobe rounded, not projecting beyond acicular lobe. Ventral ligule similar in length to acicular ligule in all chaetigers (Fig. 1F, G, I).

Notopodia homogomph spinigers only. Neuropodial heterogomph spinigers present throughout. At chaetiger 10, lower neurochaetae all heterogomph falcigers, upper neurochaetae with heterogomph falcigers, and heterogomph spinigers. At chaetiger 30 and following chaetigers, lower neurochaetae present heterogomph spinigers.

Remarks. *Perinereis wilsoni* was established by Glasby and Hsieh (2006), who provided a comprehensive description and discussion. In this study, we initially identified our specimens as *P. wilsoni* based on the key and description provided by Glasby and Hsieh (2006). Specifically, in *P. wilsoni*, the dorsal cirri are approximately 1.5 times the length of the dorsal notopodial lobe anteriorly and 2–3 times longer posteriorly; Area IV has 29.8 (± 3.6) conical paragnaths, and Area V has 1–3 conical paragnaths, usually in a longitudinal line. In contrast, in *P. mictodonta*, the dorsal cirri are either equal to or only slightly longer than the dorsal notopodial lobe throughout the body; Area IV has 35.3 (± 6.8) conical paragnaths, and Area V has 3 conical paraganths, usually in a triangle arrangement. Our specimens more closely resemble *P. wilsoni*, having dorsal cirri about 1.5 times the length of the dorsal notopodial lobe throughout the body (or only increasing slightly in length posteriorly), Area IV with 18–22 conical paragnaths, and Area V with 1 conical paragnath. However, despite the statistically significant difference in these morphometric characters (see pp. 573, Glasby and Hsieh 2006), the characters all show overlap (see pp. 560 & 572, Glasby and Hsieh 2006), and thus, morphology alone may not be sufficient to distinguish all specimens belonging to *P. wilsoni* and *P. mictodonta*.

The *ITS* genes may presently be the most effective and reliable method for accurately identifying these species, as these sequences were generated from the paratypes of *P. wilsoni* (Chen et al. 2002; Glasby and Hsieh 2006). DNA sequence-based NCBI BLAST and phylogenetic analyses of *ITS* sequences also support our identification (see next sections).

Distribution. Japan; China (based on *ITS* genes). Other records require validation using at least *ITS* data.

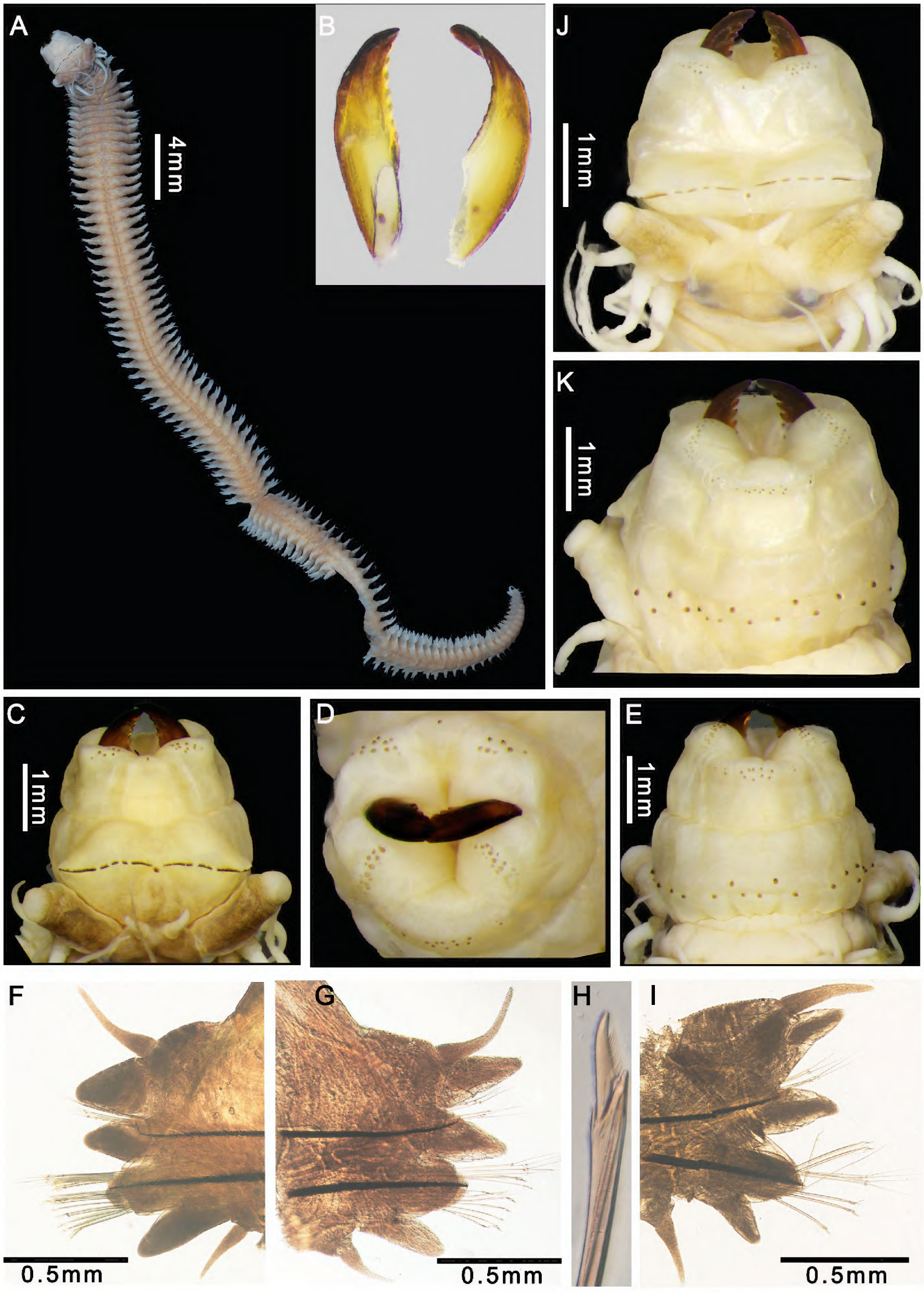

Figure 1. *Perinereis wilsoni* Glasby & Hsieh, 2006; **A–I.** (except **J, K)** 23007-1; **J, K.** 23007-2; **A.** Entire body in dorsal view; **B.** Right jaw, ventral, and dorsal view; **C.** Anterior region with pharynx everted, dorsal view; **D.** Maxillary ring, frontal view; **E.** Anterior end with pharynx everted, ventral view; **F.** Left parapodium, posterior view, chaetiger 10; **G.** Right parapodium, posterior view, chaetiger 30; **H.** Sub-acicular neuropodial heterogomph falciger, chaetiger 30; **I.** Right parapodium, posterior view, chaetiger 80; **J.** Anterior view with pharynx everted, dorsal view; **K.** Anterior view with pharynx everted, ventral view. All photos were taken by Deyuan Yang.

Characteristics of the *Perinereis wilsoni* mitochondrial genome and nuclear genes

The mitogenomes (23007-1 and 23007-2) generated from various assemblers yielded consistent results, except for the Ray assembler. They are 15,817 bp with an average coverage depth of 315×, and read mapping is 0.09% (Suppl. material 2). The genomes contain 13 protein-coding genes (PCGs), two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes, and one putative control region measuring 1160 bp. All genes are distributed on the heavy (H-) strand, similar to other Nereidinae species (Fig. 2A, Suppl. material 3: table S8). The nucleotide identity of 13 PCGs between *P. wilsoni* and the other 31 Nereididae species showed that the other species of *Perinereis* and *Neanthes acuminata* (OQ729916) have a high nucleotide identity with *P. wilsoni* (Fig. 2D).

The NCBI BLAST results of the mitogenome revealed that the sequence identity with the published *16S* sequences of *P. wilsoni* (LC482171–LC482183, Tosuji et al. 2019) ranged from 98.34% to 99.52%. However, the NCBI BLAST results for the full-length *16S* indicated over 98% sequence identity with both *P. mictodonta* (e.g., LC482161–LC482168, Tosuji et al. 2019) and *P. wilsoni* (e.g., LC482171–LC482183, Tosuji et al. 2019). These results show that the *16S* marker is not suitable for distinguishing these two species.

The nuclear rRNA contigs for the two specimens (23007-1 and 23007-2) yielded 10,983 bp and 10,820 bp with average coverages of 218.3× and 306×, respectively. Both results contained the full *18S*, *ITS1*, *5.8S*, *ITS2*, and *28S* regions, with gene lengths of 1,849 bp, 358 bp, 160 bp, 305 bp, and 3,840 bp, respectively (Fig. 2B). The nuclear histone genes for both specimens were 6,399 bp, with coverage depths of 1,547.5× and 2,620×, respectively, incorporating complete sequences of *H3* (411 bp), *H2A* (375 bp), *H2B* (372 bp), and *H4* (312 bp) genes (Fig. 2C).

The NCBI BLAST analysis of nuclear rRNA and histone genes showed that the *18S* sequence had over 99% sequence identity with eight species belonging to five genera, indicating that *18S* may not be suitable for species identification: 91.43% for *28S* with *Alitta virens* (OW028578); 98.48%-98.54% for *ITS* with *P. wilsoni* (AF332158–AF332162, Chen et al. 2002); 96%-98% for *H3* genes with *Perinereis* sp. A ZW-2022 (OL546356), *P. aibuhitensis* (MW622055), and *P. suluana* (JX443591); approximately 92% for *H2A* and *H2B* with *A. virens* (OW028584), *P. dumerilii* (X53330); and 89.4% for *H4* (Suppl. material 3: table S1).

Mitochondrial genes and codon usage

The mitochondrial genes in *P. wilsoni* exhibit a high A + T content of 64.6% (34.3% T and 30.3% A) and lower levels of C and G at 21.6% and 13.8%, respectively (Suppl. material 3: table S8). AT-skew and GC-skew are negative for the mitogenome (-0.062 and -0.221) and PCGs except for *COX2* (-0.084 and -0.250), and AT-skew is positive for tRNAs and rRNAs (0.041 and 0.043) (Suppl. material 3: table S8). The mitogenome is compact, with a total of three gene overlaps ranging in length from 2 to 7 bp. There are 12 gene spacers (259 bp in total) from 1 bp to 78 bp. All 13 PCGs start with the ATG codon and stop with T, TAA, or TAG (Suppl. material 3: table S9). A total of 22 tRNA genes with lengths ranging from 53 to 68 bp were identified in the mitogenome of *P. wilsoni*.

The most frequently utilized amino acids in the mitogenome of *P. wilsoni* are *Leu* (14.83%), *Ile* (9.94%), *Ser* (8.60%), and *Ala* (7.84%). The least common amino acids are *Cys* (1.01%), *Arg* (1.78%), *Asp* (1.78%), and *Gln* (1.94%) (Fig. 3, Suppl. material 3: table S10). Relative synonymous codon usage (RSCU) values for the 13 PCGs showed that UCU (*Ser*) and UUA (*Leu*) are the two most frequent codons, whereas UGG (*Trp*) and CCG (*Pro*) have the lowest frequencies (Fig. 3).

Phylogenetic analyses and genetic distances

Summary statistics for multiple alignments of various datasets are available in Suppl. material 3: table S11. The heterogeneity analysis of five mitogenome datasets revealed that datasets excluding the third codon position (13PCGs12, 13PCGs12_2R) exhibited lower heterogeneity compared to those including the third codon position (13PCGs, 13PCGs_2R). The lowest heterogeneity was observed in the 13PCGsAA dataset (see Suppl. material 5).

A total of ten ML and BI trees were inferred from five mitogenome datasets of Nereididae (see Suppl. material 6). These trees exhibit eight different topologies based on pairwise tree comparisons, indicating the significant influence of the chosen mitogenomic dataset and tree-building method on the inferred phylogenetic relationships. In general, ML and BI trees inferred from the same dataset are generally consistent, except for some obvious differences observed in the 13PCGs_2R and 13PCGs12_2R datasets (see Suppl. material 6). In the 13PCGs12 dataset, we could not re-root the ML and BI trees using the outgroup, which was excluded from TreeSpace analysis. The remaining eight trees were clustered into four groups, representing four types of trees (see Fig. 4B–F). The most common topology is depicted in Fig. 4A, D.

Phylogenetic analyses based on five datasets of the available mitochondrial genomes (32 taxa in Nereididae, including 8 taxa in *Perinereis*) provide support for the monophyly of the genus *Perinereis*. *Perinereis* was either sister to the genus *Nereis* and the species *Cheilonereis cyclurus* (MF538532) (Fig. 4C, E, F) or sister to the genera *Platynereis*, *Hediste*, *Alitta*, *Nectoneanthes*, and the species *Laeonereis culveri* (KU992689) (Fig. 4D). In all trees, *P. wilsoni* was sister to *P. vancaurica* and *P. nuntia* with high nodal support values (BS ≥ 94%, PP = 1) (Fig. 4 and Suppl. material 6). The *P. nuntia* species group was found not to be monophyletic. Specifically, *P. linea* (NC 063944), *P. aibuhitensis* (NC 023943), and *P. vancaurica* (ON611802) have two or three paragnaths on Area VI of the pharynx,

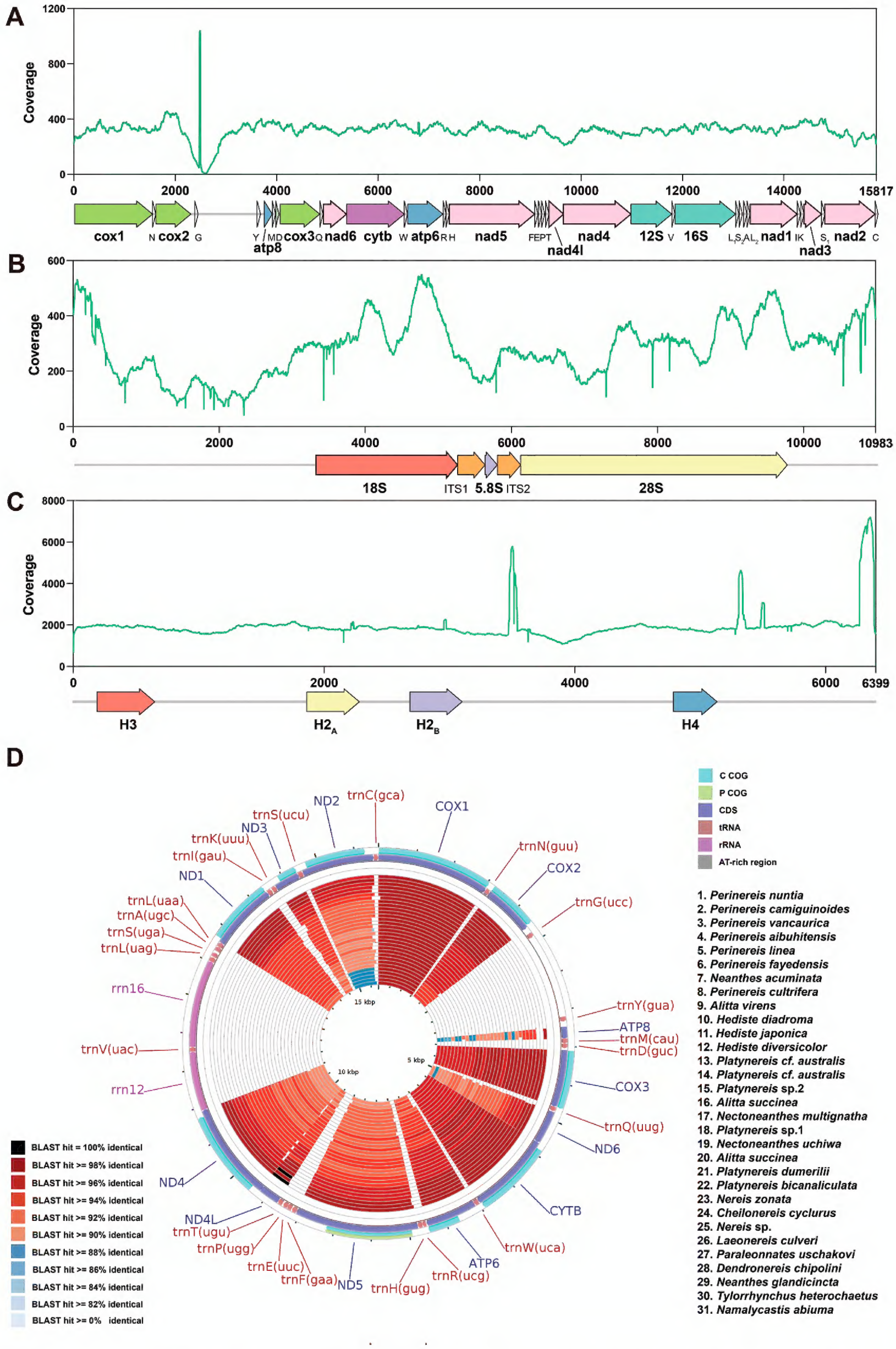

Figure 2. The gene map of *Perinereis wilsoni* (23007-1), 23007-2, is the same as 23007-1 but only shows one; **A.** Mitogenome; **B.** Nuclear rRNA cluster; **C.** And histone genes; **D.** CCT (CGView Comparison Tool) map and sequence identity compare the mitogenome between *P. wilsoni* and the other nereidids. **A–C.** The green lines depict the distribution of coverage depth; **D.** Starting from the outermost ring, the feature rings depict: 1. COG (Clusters of Orthologous Groups of Proteins) functional categories for forward strand coding sequences; 2. Forward strand sequence features; 3. The remaining rings show regions of sequence similarity detected by BLAST comparisons between CDS translations from the reference genome and 31 comparison genomes. BLAST identities are organized from high to low, with higher values closer to the outer ring.

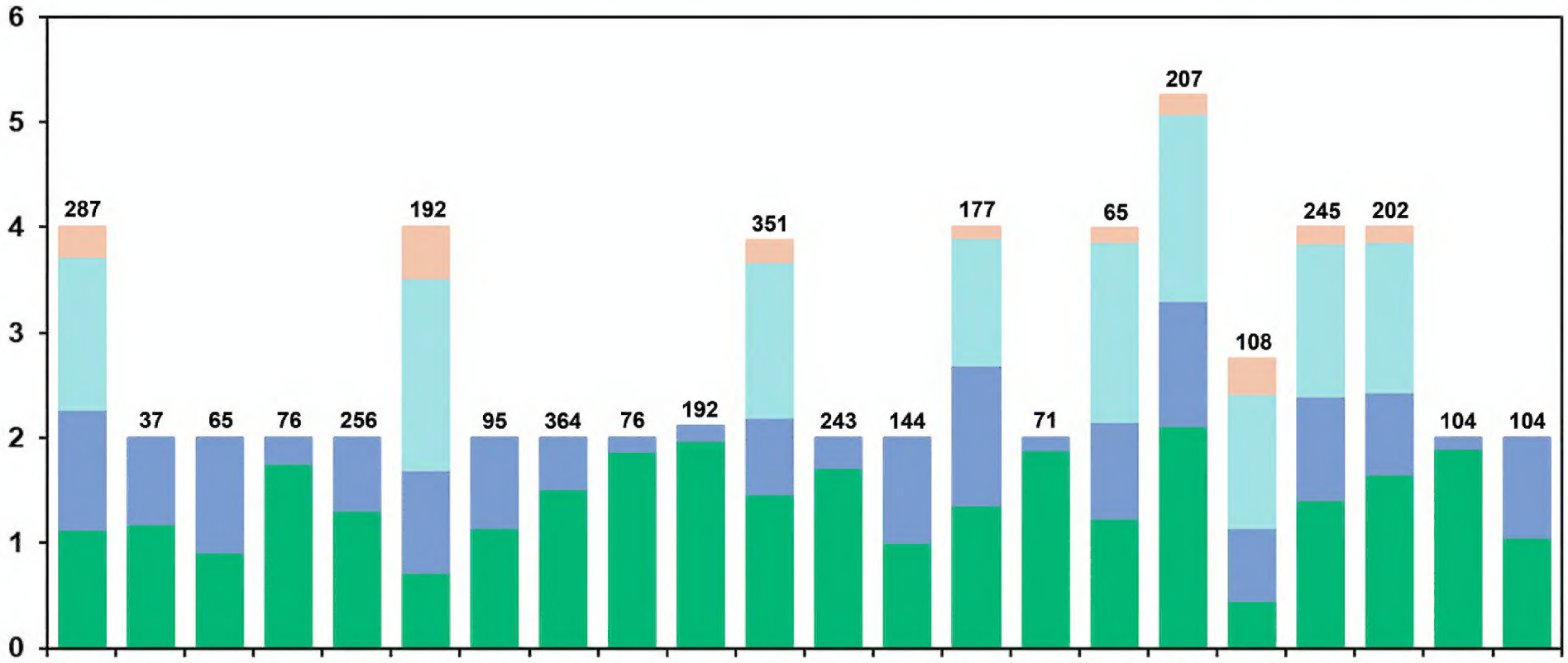

Figure 3. The relative synonymous codon usage (RSCU) in the mitogenome of *Perinereis wilsoni*. The numbers on the bars represent amino acid composition.

which belong to *Perinereis* Group 2 (Hutchings et al. 1991); all other *Perinereis* species (except *Perinereis cultrifera*) in the trees belong to the *P. nuntia* group. Yet, the species of Group 2 were nested within the *P. nuntia* group.

Phylogenetic trees (ML and BI) based on *COX1* genes recovered that *P. wilsoni* (23007-1, 23007-2) have a closer relationship with *P. mictodonta* (KC800632, KC800630, KC800628), with nodal support values (BS = 83%, PP = 0.94). All taxa labeled *P. wilsoni* and *P. mictodonta* did not each form a monophyletic group, respectively; instead, they were divided into five distinct clades. This suggests the potential of cryptic species present within *P. wilsoni* and *P. mictodonta* or that the specimens were sampled from geographically distant localities with some degree of isolation. These five clades were also supported by genetic distances, which have a distinct barcode gap from each other (Fig. 5A, D, and Suppl. material 7). The phylogenetic trees (ML and BI) based on *16S* genes suggested that *P. wilsoni* and *P. mictodonta* were each not monophyletic (Fig. 5B, E, and Suppl. material 8). Genetic distances also showed that the sequences of *P. wilsoni* and *P. mictodonta* had no barcode gap (Fig. 5B). The phylogenetic trees (ML and BI) based on *ITS* genes supported *P. wilsoni* as a sister to *P. mictodonta* (BS = 96%, PP = 1). Although *P. wilsoni* and *P. mictodonta* formed a monophyletic cluster with each other, the genetic distance analyses showed that these two species have no distinct barcode gap (Fig. 5C, F, and Suppl. material 9). Based on the dos Santos Vieira et al. (2020) methodology, optimal markers were selected based on an overlap criterion of less than 20%. However, the overlap between *P. wilsoni* and *P. mictodonta* exceeded 35%, indicating that *COX1*, *16S*, and *ITS* genes are not optimal molecular markers (Fig. 5A–C).

The positions of *Paraleonnates uschakovi* (NC 032361) and *Laeonereis culveri* (KU992689) are unstable, jumping across different phylogenies (Fig. 4, Suppl. material 6). *Laeonereis culveri* (KU992689) is closer to *P. uschakovi* (NC 032361) in four trees (13PCGs_2R_ML, 13PCGs12_ML, 13PCGs12_2R_ML, and 13PCGsAA_BI in Suppl. material 6), all with high nodal support. In contrast, in six other trees, *L. culveri* (KU992689) is more closely related to the genera *Hediste*, *Alitta*, and *Nectoneanthes*, also with high nodal support, except in the 13PCGs12_ML and 13PCGs12_BI trees (Fig. 4, Suppl. material 6). *Paraleonnates uschakovi* is closer to the subfamilies Gymnonereidinae, Dendronereinae, and Nereidinae (8 of 10 trees), with low nodal support in ML trees but with high nodal support in BI trees. In 13PCGs12_ML and 13PCGs12_BI trees, *P. uschakovi* was clustered with the outgroup, all with low nodal support (Suppl. material 6).

There are two primary types of mitochondrial gene order in the known mitochondrial genomes of Nereididae, except for *L. culveri* (KU992689). The first type of gene order is observed in the subfamilies Gymnonereidinae, Dendronereinae, and *P. uschakovi*. The second type is found in the subfamily Nereidinae, except for *L. culveri* (see Fig. 4).

Nucleotide diversity and evolutionary rate analyses

The nucleotide diversity (Pi) analysis was conducted using concatenated alignments of 13 PCGs and

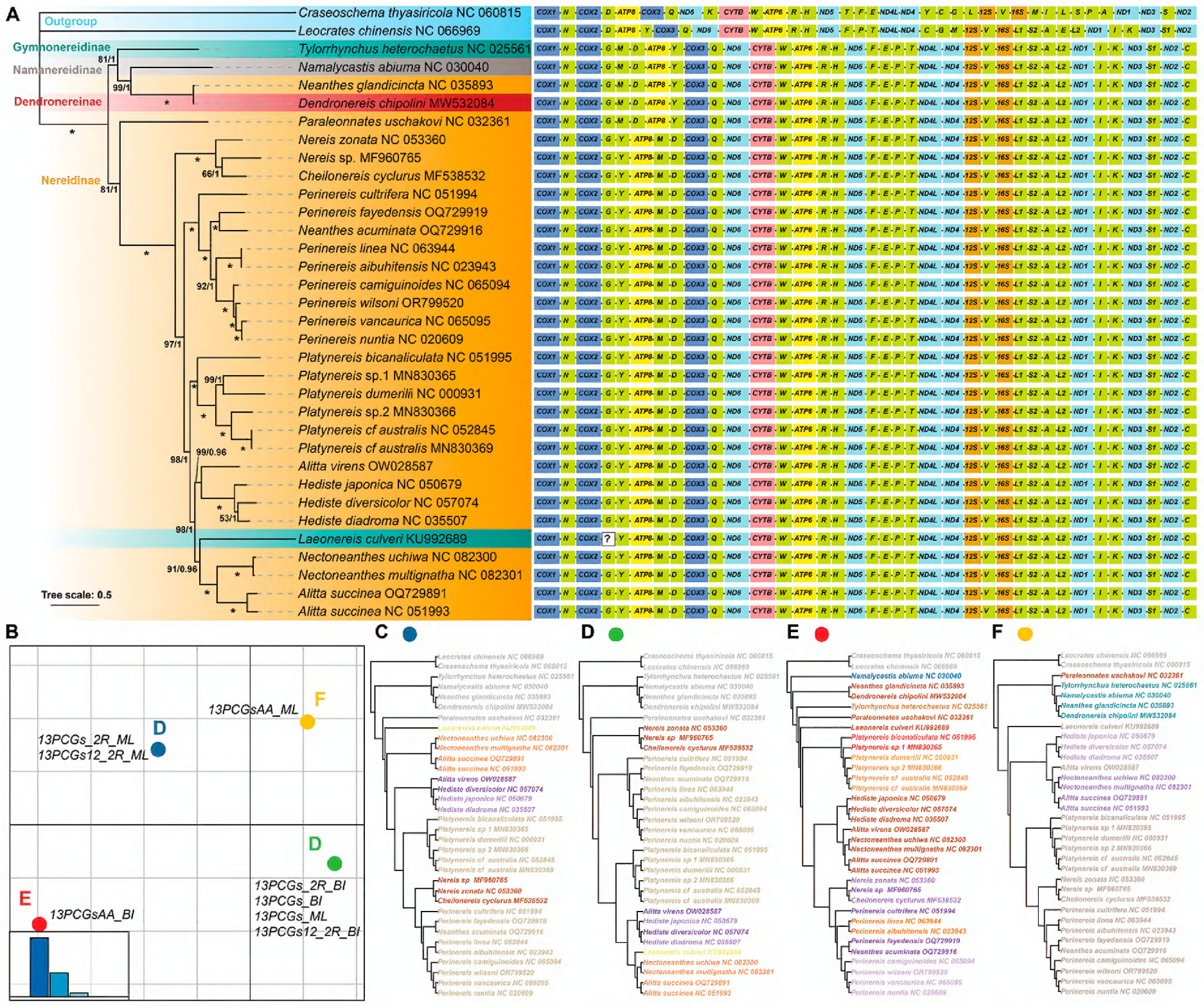

Figure 4. An analysis of phylogenies from five mitogenome datasets. **A.** Maximum likelihood (ML) and Bayesian inference (BI) tree of Nereididae based on the dataset 13PCGs. The GenBank accession numbers used are listed after the species names. The scale bar (0.5) corresponds to the estimated number of substitutions per site. Numbers at nodes are statistical support values for ML bootstrap support. Asterisks denote 100% bootstrap support. "-" indicates no support value. Color-coded clades are four subfamilies within Nereididae. The gene order is shown to the right. "?" indicates the deletion of a gene; **B.** A two-dimensional MDS plot of eight trees (excluding 13PCGs12_ML and 13PCGs12_BI), colored by different clusters. **C–F.** For each cluster identified in **B**., a representative tree was selected. **Note:** *Neanthes glandicincta* (NC 035893) is an incorrectly identified taxon, which should be the genus *Dendronereis*. *Neanthes acuminata* (OQ729916) and *Perinereis fayedensis* (OQ729919) should be *Perinereis suezensis* and *Perinereis damietta*, respectively. *Perinereis aibuhitensis* (NC 023943) should be *Perinereis linea* (NC 063944).

two rRNAs of 32 Nereididae species. The sequence variation ratio exhibits variable nucleotide diversity between the Nereididae, with Pi values for the 100 bp windows ranging from 0.061 to 0.563 (Fig. 6A). *COX1* (Pi = 0.211), *COX3* (0.235), *CYTB* (0.238), and *COX2* (0.239) exhibit a comparatively low sequence variability, whereas *ATP8* (0.405), *ND2* (0.386), and *ND6* (0.378) have a comparatively high sequence variability. This is corroborated by the non-synonymous/synonymous (Ka/Ks) ratio analysis, which shows that *COX1* (Ka/Ks = 0.036), *COX3* (0.038), *CYTB* (0.053), and *COX2* (0.065) are evolving comparatively slowly, whereas *ATP8* (0.270), *ND2* (0.196), and *ND6* (0.173) are evolving comparatively fast (Fig. 6B). All genes are under purifying selection.

Discussion

The taxonomic status of *Perinereis wilsoni*

Before discussing this topic, it is crucial to specify the species concept employed in this study. Despite approximately 30 species concepts having been proposed (Hong 2020), the scientific community has not reached a consensus on defining a species. Our study adopts the Gen-morph species concept, initially proposed by Deyuan Hong (Hong 2020), which requires at least two each of independent morphological characteristics and genetic markers for species definition. According to this concept, morphological characters include quantitative morphological features that have been shown to be statistically different between species.

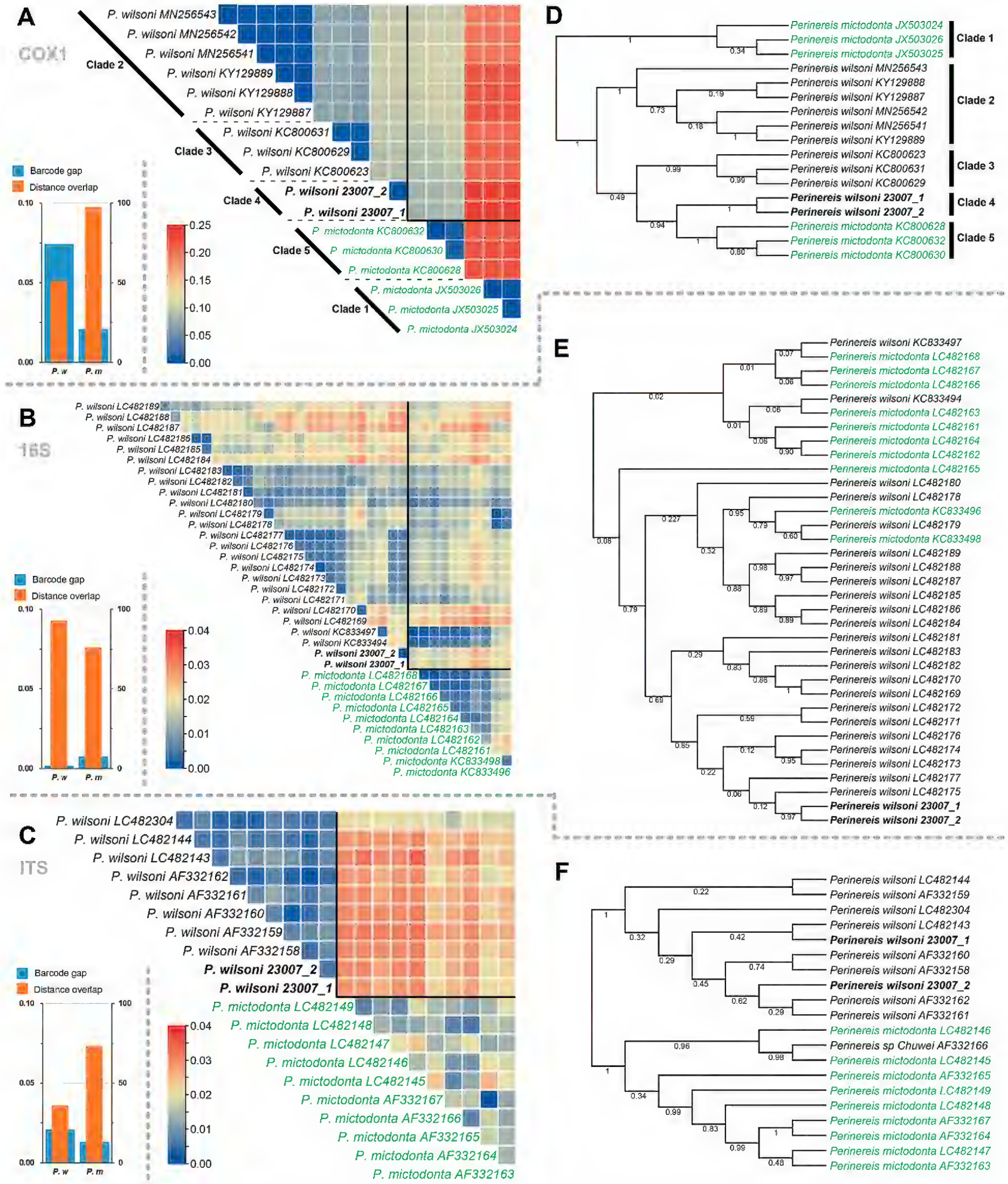

Figure 5. The heatmap of *COX1* **A**, *16S* **B.**, and *ITS* **C.** p-distance for *Perinereis wilsoni* and *P. mictodonta*, with the barcode gap and distance overlap of them on the left. Bayesian inference (BI) phylogenetic trees for the two species, based on sequences of *COX1* **D**, *16S* **E**, and *ITS* **F**, are excerpted from Suppl. materials 7–9. Sequences from *P. wilsoni* are depicted in black, those from *P. mictodonta* in green, with sequences from this study highlighted in bold.

Perinereis wilsoni and *P. mictodonta* were initially established as separate species based on statistically validated morphometric differences and the results of the *ITS* genes (Glasby and Hsieh 2006). Subsequent studies by Park and Kim (2007) and Tosuji et al. (2019) demonstrated that *COX1*, *16S*, and *ITS* genes can distinguish these two species, respectively. However, the *COX1* sequences used in Park and Kim (2007) have not been made publicly available. Moreover, their study relied solely on morphological characteristics for species identification and did not validate these identifications with *ITS* gene analysis, casting doubt on the reliability of their findings. Our re-analysis of sequences from (Tosuji et al. 2019) showed that the *16S* was insufficient to distinguish the

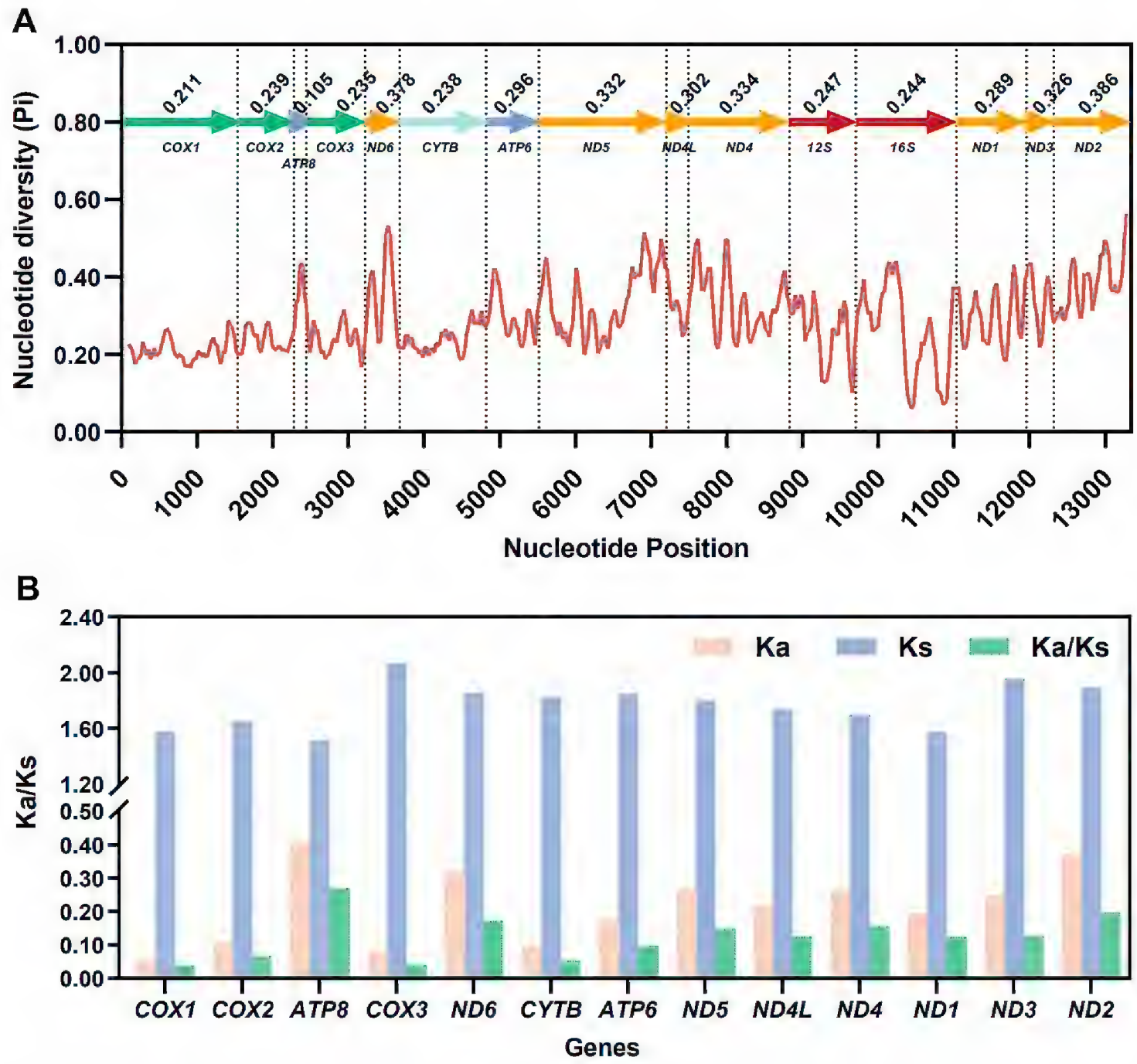

Figure 6. Nucleotide diversity analysis: **A.** Of 13 PCGs + two rRNAs and Ka/Ks rates; **B.** Of 13 PCGs based on 32 Nereididae species. The Pi values for the 13 PCGs + two rRNAs are shown on the graph. The red line represents the value of nucleotide diversity (Pi) (window size = 100 bp, step size = 20 bp). The pink, purple, and green columns represent the values of Ka, Ks, and Ka/Ks, respectively.

two species. To date, only *ITS* genes have proven effective in distinguishing *P. wilsoni* and *P. mictodonta* (Chen et al. 2002; Glasby and Hsieh 2006; Tosuji et al. 2019). Considering the Gen-morph species concept, *P. wilsoni* may not be a "good" species until a second genetic marker (and a more optimal one than *ITS*; see below) can be found. Therefore, comprehensive specimen sampling from various geographical locations is still required to assess its taxonomic status thoroughly (Deyuan Yang et al. in preparation).

Which molecular markers suit *Perinereis wilsoni* Glasby & Hsieh, 2006 identification?

Partial mitochondrial genes, like *COX1* and *16S*, and nuclear genes, like *18S*, *28S*, *ITS*, and *H3*, are widely used in Nereididae for species discovery and phylogenetic studies (Glasby et al. 2013; Elgetany et al. 2022; Teixeira et al. 2022a; Teixeira et al. 2022b; Teixeira et al. 2024). Among these genes, the *18S* and *28S* genes are typically used in higher-taxon phylogenetic analyses. A comprehensive assessment of the effectiveness of these molecular markers in differentiating species, both closely and distantly related across various genera and families, remains limited (Halanych and Janosik 2006).

Given that species within the *P. nuntia* complex are morphologically similar and not easily distinguishable based on morphology alone, molecular-based identification can offer a faster and more reliable method for species identification when reliable molecular references are available. Our analyses were unable to confirm the discriminative capability of the *COX1* gene between *P. wilsoni* and *P. mictodonta*, as the sequences sourced from public databases were not linked to morphological and *ITS* gene data, thus limiting our further study. Additionally, the *16S* gene was proven to be insufficient to distinguish these two species. Although *ITS* genes have been found effective in distinguishing cryptic species (Chen et al. 2002; Pleijel et al. 2009; Nygren and Pleijel 2011; Nygren 2014), they may not be the optimal molecular marker in distinguishing *P. wilsoni* and *P. mictodonta* due to the lack of a distinctive barcoding gap between these two species. Furthermore, there are only a few available sequences of Polychaeta *ITS* genes in the public database, possibly due to difficulties in amplifying these sequences by PCR. We found that the primer sets *ITS18SFPOLY* and *ITS28SRPOLY* (Nygren et al. 2009) may be more effective than the one provided by Chen et al. (2002) for amplifying *ITS* genes in *Perinereis*.

With advances in sequencing technologies, high-throughput sequencing (HTS) is more efficient than PCR-based Sanger sequencing in obtaining molecular markers. Recently,

some new molecular markers have been proposed, such as nearly universal single-copy nuclear protein-coding genes (Eberle et al. 2020) and organelle genomes (Margaryan et al. 2021). In this paper, we found the genome-skimming approach makes it easier to explore the mitogenome, complete nuclear ribosomal DNA (*18S-ITS1-5.8S-ITS2-28S*), and complete histone genes. However, it is more difficult to explore universal single-copy nuclear protein-coding genes, even with sequencing data increased to 25 Gb.

The phylogeny of *Perinereis* and the systematic position of *Perinereis wilsoni*

Currently, a robust phylogenetic backbone of the genus *Perinereis*, based on phylogenomic methods and extensive taxon sampling encompassing major species in all five informal grouping schemes proposed by Hutchings et al. (1991), is still lacking. Traditionally, the genus *Perinereis* has been considered polyphyletic (Bakken and Wilson 2005; Glasby et al. 2013; Alves et al. 2020; Elgetany et al. 2022). However, this view is largely based on morphological evidence or a limited number of loci and/or may include assembly errors in their datasets. In this study, although our curated mitogenome datasets support the monophyly of the genus *Perinereis*, the dataset includes only 8 species, whereas the genus currently comprises 103 species (Prajapat et al. 2024). Furthermore, morphological diversity within the genus has been unevenly sampled. Hutchings et al. (1991) proposed an informal grouping scheme with five groups. Our mitogenome dataset includes Group 1A, represented by 1 taxon (*P. cultrifera*); Group 2A, represented by 4 taxa (*P. linea*, *P. aibuhitensis* (should be *P. linea*), *P. camiguinoides*, and *P. vancuarica*); and Group 3A, represented by 4 taxa (*P. fayedensis*, *N. acuminata* (=*P. suezensis*), *P. wilsoni*, and *P. nuntia*). However, Groups 1B and 3B are not represented. To confidently establish the monophyly of the genus, future analyses should include multiple representatives from each of these informal morphological groupings.

Based on the available mitogenome datasets, *Perinereis wilsoni* is a sister group to *P. vancaurica* and *P. nuntia*, with high nodal support in all phylogenetic trees. In contrast, single-gene phylogenetic trees suggest that *P. wilsoni* is more closely related to *P. mictodonta*, with low nodal support. Although including more taxa, single-gene trees do not provide sufficient resolution in phylogeny. In summary, the phylogenetic relationships of *P. wilsoni* remain poorly understood due to the limited number of *Perinereis* species or other Nereididae for which genomic data are available.

The phylogeny of Nereididae

A deep discussion of the phylogeny of Nereididae is beyond the scope of our current work. Here, we provide a brief discussion. The positions of *P. uschakovi* and *L. culveri* (KU992689) were observed to be unstable in the trees, inferred from different mitogenome datasets in this study.

In detail, *L. culveri* always nested within the subfamily Nereidinae, which was also found in previous studies using different datasets, such as *COX1*, *16S*, and *18S* (Wang et al. 2021; Alves et al. 2023; Villalobos-Guerrero et al. 2024), as well as mitogenome datasets (Alves et al. 2020). *L. culveri* is also a sister group to the Gymnonereidinae, Dendronereinae, and Nereidinae subfamilies, along with *P. uschakovi* (this study; Villalobos-Guerrero et al. 2024). *P. uschakovi* was always found as a sister group to other subfamilies (Alves et al. 2020; Wang et al. 2021; Alves et al. 2023; Villalobos-Guerrero et al. 2024; this study). However, it occasionally grouped with the outgroup *Craseoschema thyasiricola* (NC 060815, see Suppl. material 6: 13PCGs12_ML tree and 13PCGs12_BI tree) or formed a single clade with the Gymnonereidinae and Dendronereinae subfamilies (see Suppl. material 6: 13PCGsAA_ML tree), which appears to be supported by mitochondrial genome gene order. Alves et al. (2020) also found this situation when removing outgroups. Inspired by a morphological phylogenetic tree proposed by Wu et al. (1981), which is based on the characteristics of pharyngeal armature, we propose a hypothesis that *Paraleonnates* and *Laeonereis* (KU992689) may represent early branching members of Nereidinae. The lack of sufficient data on such taxa may be causing their uncertain placements within the phylogenetic framework.

In this study, we uncovered potential errors in assembly and annotation within GenBank. However, the absence of corresponding Sequence Read Archive (SRA) data in public databases hampers accurate confirmation of these errors. Consequently, we filtered these sequences based solely on our expertise. Therefore, I (Deyuan Yang) advocate for the uploading of original data (raw data) to public databases, such as NCBI. Even if some authors are unwilling to upload their data, various assembly methods should be employed to ensure the accuracy of the assemblies.

Additionally, we emphasize that carefully curated datasets (verifying the taxonomic identification of the species used in the phylogenetic study), especially those from public databases, are crucial before conducting phylogenetic studies, as taxonomic misidentifications can lead to incorrect conclusions. For example, Alves et al. (2020) concluded that Gymnonereidinae was non-monophyletic, a finding that was impacted by the inclusion of an incorrectly identified taxon, *Neanthes glandicincta* (NC 035893), which should have been classified under the genus *Dendronereis* Peters, 1854 (Zhen et al. 2022; this study). Additionally, in this study, if we had not questioned the name *Neanthes acuminata* (OQ729916), we would have concluded that *Perinereis* is non-monophyletic. However, *N. acuminata* should be re-identified as *Perinereis suezensis* (Elgetany et al. 2022).

Author Contribution

D.Y.Y. and S.Z. wrote the manuscript, and S.Z. wrote the part on mitochondrial genes, codon usage, nucleotide diversity, and evolutionary rate analyses. D.Y.Y. and S.Z.

analyzed the data. Z.W., Y.J.Z., and D.Z.Y. participated in the discussion and reviewed the manuscript. C.J.G., J.S.H., and L.Z.C. conceived and designed, supervised the work, and reviewed drafts of the paper. S.Z. and D.Y.Y. contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Many thanks to Yuanzheng Meng for helping us organize and format the literature for this paper. Thanks to Grammarly (https://www.grammarly.com/) for grammar correction while writing the first manuscript and ChatGPT 4.0 for generating some Python and R scripts during this study. Thanks to reviewer Robin Wilson for his constructive comments, especially focused on the phylogeny of the genus *Perinereis*, which have greatly contributed to the revision of our article. This work was supported by the Youth Fund of the National Natural Science Foundation of China (42306107) and the China Postdoctoral Science Foundation (2021M691866).

References

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F (2020) MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Molecular Ecology Resources 20(4): 892–905. https://doi.org/10.1111/1755-0998.13160

Alves PR, Halanych KM, Santos CSG (2020) The phylogeny of Nereididae (Annelida) based on mitochondrial genomes. Zoologica Scripta 49(3): 366–378. https://doi.org/10.1111/zsc.12413

Alves PR, Halanych KM, Silva EP, Santos CS (2023) Nereididae (Annelida) phylogeny based on molecular data. Organisms Diversity & Evolution, 1–13. https://doi.org/10.1007/s13127-023-00608-9

Arias A, Richter A, Anadón N, Glasby CJ (2013) Revealing polychaetes invasion patterns: Identification, reproduction and potential risks of the Korean ragworm, *Perinereis linea* (Treadwell), in the Western Mediterranean. Estuarine, Coastal and Shelf Science 131: 117–128. https://doi.org/10.1016/j.ecss.2013.08.017

Bakken T, Wilson RS (2005) Phylogeny of nereidids (Polychaeta, Nereididae) with paragnaths. Zoologica Scripta 34(5): 507–547. https://doi.org/10.1111/j.1463-6409.2005.00200.x

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455–477. https://doi.org/10.1089/cmb.2012.0021

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. Journal of Computational Biology 17(11): 1519–1533. https://doi.org/10.1089/cmb.2009.0238

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15): 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen CA, Chen C-P, Fan T-Y, Yu J-K, Hsieh H-L (2002) Nucleotide sequences of ribosomal internal transcribed spacers and their utility in distinguishing closely related *Perinereis* polychaets (Annelida; Polychaeta; Nereididae). Marine Biotechnology 4: 17–29. https://doi.org/10.1007/s10126-001-0069-3

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17): i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 13(8): 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: extending the concept of DNA barcoding. Molecular Ecology 25(7): 1423–1428. https://doi.org/10.1111/mec.13549

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45(4): e18-e18. https://doi.org/10.1093/nar/gkw955

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M (2019) Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Research 47(20): 10543–10552. https://doi.org/10.1093/nar/gkz833

dos Santos Vieira WA, Bezerra PA, da Silva AC, Veloso JS, Câmara MPS, Doyle VP (2020) Optimal markers for the identification of Colletotrichum species. Molecular Phylogenetics and Evolution 143: 106694. https://doi.org/10.1016/j.ympev.2019.106694

Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B (2020) A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends in Ecology & Evolution 35(4): 336–345. https://doi.org/10.1016/j.tree.2019.12.003

Elgetany AH, Struck TH, Glasby CJ (2022) Three new species of the genus *Perinereis* (Annelida, Nereididae) from Egyptian coasts. ZooKeys 1132: 163–188. https://doi.org/10.3897/zookeys.1132.87629

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32(19): 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Glasby CJ (1993) Family revision and cladistic analysis of the Nereidoidea (Polychaeta: Phyllodocida). Invertebrate Systematics 7(6): 1551–1573. https://doi.org/10.1071/IT9931551

Glasby CJ, Hsieh H-L (2006) New species and new records of the *Perinereis nuntia* species group (Nereididae: Polychaeta) from Taiwan and other Indo-West Pacific shores. Zoological Studies 45(4): 553–577. https://zoolstud.sinica.edu.tw/Journals/45.4/553.pdf

Glasby CJ, Wei N-WV, Gibb KS (2013) Cryptic species of Nereididae (Annelida: Polychaeta) on Australian coral reefs. Invertebrate Systematics 27(3): 245–264. https://doi.org/10.1071/IS12031

Grant JR, Arantes AS, Stothard P (2012) Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics 13: 1–8. https://doi.org/10.1186/1471-2164-13-202

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8): 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41(13): e129-e129. https://doi.org/10.1093/nar/gkt371

Halanych KM, Janosik AM (2006) A review of molecular markers used for Annelid phylogenetics. Integrative and Comparative Biology 46(4): 533–543. https://doi.org/10.1093/icb/icj052

Hektoen MM, Bakken T, Ekrem T, Radashevsky VI, Dunshea G (2024) Species delimitation and phylogenetic relationships of the *Prionospio* complex (Annelida, Spionidae) in the Northeast Atlantic. Zoologica Scripta: 358–375. https://doi.org/10.1111/zsc.12648

Hong DY (2020) Gen-morph species concept—A new and integrative species concept for outbreeding organisms. Journal of systematics and evolution 58(5): 725–742. https://doi.org/10.1111/jse.12660

Hutchings P, Lavesque N (2020) I know who you are, but do others know? Why correct scientific names are so important for the biological sciences. Zoosymposia 19: 151–163. https://doi.org/10.11646/zoosymposia.19.1.16

Hutchings P, Reid A, Wilson R (1991) *Perinereis* (Polychaeta, Nereididae) from Australia, with redescriptions of six additional species. Records of the Australian Museum 43(3): 241–274. https://doi.org/10.3853/j.0067-1975.43.1991.47

Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology 21: 1–31. https://doi.org/10.1186/s13059-020-02154-5

Jombart T, Kendall M, Almagro-Garcia J, Colijn C (2017) treespace: Statistical exploration of landscapes of phylogenetic trees. Molecular Ecology Resources 17(6): 1385–1392. https://doi.org/10.1111/1755-0998.12676

Köster J, Rahmann S (2012) Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28(19): 2520–2522. https://doi.org/10.1093/bioinformatics/bts480

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35(6): 1547. https://doi.org/10.1093/molbev/msy096

Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49(W1): W293-W296. https://doi.org/10.1093/nar/gkab301

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997: https://doi.org/10.48550/arXiv.1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16): 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li D, Liu C-M, Luo R, Sadakane K, Lam T-W (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10): 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Mahcene HR, Villalobos-Guerrero TF, Kurt G, Denis F, Daas T (2023) A new species of *Perinereis* Kinberg, 1865 (Annelida: Nereididae) from the Western Mediterranean Sea revealed by morphological and molecular approaches. Mediterranean Marine Science 24(2): 454–460. https://doi.org/10.12681/mms.33969

Margaryan A, Noer CL, Richter SR, Restrup ME, Bülow-Hansen JL, Leerhøi F, Langkjær EMR, Gopalakrishnan S, Carøe C, Gilbert MTP (2021) Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database, DNAmark. Environmental DNA 3(2): 472–480. https://doi.org/10.1002/edn3.138

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71(2): 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Meng G, Li Y, Yang C, Liu S (2019) MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research 47(11): e63. https://doi.org/10.1093/nar/gkz173

Nygren A (2014) Cryptic polychaete diversity: a review. Zoologica Scripta 43(2): 172–183. https://doi.org/10.1111/zsc.12044

Nygren A, Pleijel F (2011) From one to ten in a single stroke–resolving the European *Eumida sanguinea* (Phyllodocidae, Annelida) species complex. Molecular Phylogenetics and Evolution 58(1): 132–141. https://doi.org/10.1016/j.ympev.2010.10.010

Nygren A, Eklöf J, Pleijel F (2009) Arctic-boreal sibling species of *Paranaitis* (Polychaeta, Phyllodocidae). Marine Biology Research 5(4): 315–327. https://doi.org/10.1080/17451000802441301

Palmer PJ (2010) Polychaete-assisted sand filters. Aquaculture 306(1–4): 369–377. https://doi.org/10.1016/j.aquaculture.2010.06.011

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Park T-S, Kim W (2007) A taxonomic study on *Perinereis nuntia* species group (Polychaeta: Nereididae) of Korea. Animal Systematics, Evolution and Diversity 23(1): 75–85. https://doi.org/10.5635/KJSZ.2007.23.1.075

Peng Y, Leung HC, Yiu S-M, Chin FY (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28(11): 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of molecular evolution 41: 353–358. https://doi.org/10.1007/BF01215182

Pleijel F, Dahlgren T (1998) Position and delineation of Chrysopetalidae and Hesionidae (Annelida, Polychaeta, Phyllodocida). Cladistics 14(2): 129–150. https://doi.org/10.1111/j.1096-0031.1998.tb00327.x

Pleijel F, Rouse G, Nygren A (2009) Five colour morphs and three new species of *Gyptis* (Hesionidae, Annelida) under a jetty in Edithburgh, South Australia. Zoologica Scripta 38(1): 89–99. https://doi.org/10.1111/j.1463-6409.2008.00356.x

Prajapat V, Villalobos-Guerrero TF, Vachhrajani KD (2023) A new species of *Perinereis* Kinberg, 1865 (Annelida: Nereididae) and invalidation of two congeners from Western India. Zootaxa 5330(3): 398–412. https://doi.org/10.11646/zootaxa.5330.3.4

Prajapat V, Villalobos-Guerrero TF, Vachhrajani KD (2024) Two new species of *Perinereis* Kinberg, 1865 (Annelida: Nereididae) from the rocky shore of Maharashtra, India, including notes and an identification key to Group 1. European Journal of Taxonomy 935: 256–282. https://doi.org/10.5852/ejt.2024.935.2561

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35(10): 2582–2584. https://doi.org/10.1093/molbev/msy159

Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C (2015) The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). Genome Biology and Evolution 7(12): 3443–3462. https://doi.org/10.1093/gbe/evv224

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34(12): 3299–3302. https://doi.org/10.1093/molbev/msx248

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Song M-H, Yan C, Li J-T (2022) MEANGS: an efficient seed-free tool for de novo assembling animal mitochondrial genome using whole genome NGS data. Briefings in Bioinformatics 23(1): bbab538. https://doi.org/10.1093/bib/bbab538

Steenwyk JL, Buida III TJ, Gonçalves C, Goltz DC, Morales G, Mead ME, LaBella AL, Chavez CM, Schmitz JE, Hadjifrangiskou M (2022) BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. Genetics 221(3): iyac079. https://doi.org/10.1093/genetics/iyac079

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. American Journal of Botany 99(2): 349–364. https://doi.org/10.3732/ajb.1100335

Struck TH, Purschke G, Halanych KM (2006) Phylogeny of Eunicida (Annelida) and exploring data congruence using a partition addition bootstrap alteration (PABA) approach. Systematic Biology 55(1): 1–20. https://doi.org/10.1080/10635150500354910

Teixeira MA, Bakken T, Vieira PE, Langeneck J, Sampieri BR, Kasapidis P, Ravara A, Nygren A, Costa FO (2022a) The curious and intricate case of the European *Hediste diversicolor* (Annelida, Nereididae) species complex, with description of two new species. Systematics and Biodiversity 20(1): 1–39. https://doi.org/10.1080/14772000.2022.2116124

Teixeira MA, Langeneck J, Vieira PE, Hernández JC, Sampieri BR, Kasapidis P, Mucciolo S, Bakken T, Ravara A, Nygren A (2022b) Reappraisal of the hyperdiverse *Platynereis dumerilii* (Annelida: Nereididae) species complex in the Northern Atlantic, with the description of two new species. Invertebrate Systematics 36(11): 1017–1061. https://doi.org/10.1071/IS21084

Teixeira MA, Fourreau CJL, Sampere-Valverde J, Carvalho S (2024) Two new records and description of a new *Perinereis* (Annelida, Nereididae) species for the Saudi Arabian Red Sea region. ZooKeys 1196: 331–354. https://doi.org/10.3897/zookeys.1196.115260

Tilic E, Stiller J, Campos E, Pleijel F, Rouse GW (2022) Phylogenomics resolves ambiguous relationships within *Aciculata* (Errantia, Annelida). Molecular Phylogenetics and Evolution 166: 107339. https://doi.org/10.1016/j.ympev.2021.107339

Tosuji H, Nishinosono K, Hsieh H-L, Glasby CJ, Sakaguchi T, Sato M (2019) Molecular evidence of cryptic species diversity in the *Perinereis nuntia* species group (Annelida: Nereididae) with first records of *P. nuntia* and *P. shikueii* in southern Japan. Plankton and Benthos Research 14(4): 287–302. https://doi.org/10.3800/pbr.14.287

Tosuji H, Park T, Goryo Y, Kan K, Abe H, Sato M (2023) Molecular method to identify the morphologically similar four species of the *Perinereis nuntia* species complex (Annelida: Nereididae) based on PCR-RFLP analysis of nuclear ribosomal ITS, with new distributional records of the two forms of *Perinereis shikueii*. Plankton and Benthos Research 18(3): 124–130. https://doi.org/10.3800/pbr.18.124

Villalobos-Guerrero TF (2019) Redescription of two overlooked species of the *Perinereis nuntia* complex and morphological delimitation of *P. nuntia* (Savigny in Lamarck, 1818) from the Red Sea (Annelida, Nereididae). Zoosystema 41(1): 465–496. https://doi.org/10.5252/zoosystema2019v41a24

Villalobos-Guerrero TF, Huč S, Tilic E, Hiley AS, Rouse GW (2024) A remarkable new deep-sea nereidid (Annelida: Nereididae) with gills. PLoS ONE 19(3): e0297961. https://doi.org/10.1371/journal.pone.0297961

Wang Y, Cheng H, Wang C (2021) A new eyeless species of *Nicon* (Annelida: Nereididae) from the deep Northwest Pacific Ocean. Acta Oceanologica Sinica 40(12): 20–26. https://doi.org/10.1007/s13131-021-1886-z

Wickham H (2016) Data Analysis. In: ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing, Cham, 189–201. https://doi.org/10.1007/978-3-319-24277-4_9

Wilson RS, Glasby CJ (1993) A revision of the *Perinereis nuntia* species group (Polychaeta: Nereididae). Records of the Australian Museum 45(3): 253–277. https://doi.org/10.3853/j.0067-1975.45.1993.23

Wilson RS, Glasby CJ, Bakken T (2023) The Nereididae (Annelida)–diagnoses, descriptions, and a key to the genera. ZooKeys 1182: 35–134. https://doi.org/10.3897/zookeys.1182.104258

Wu B, Sun R, Yang D (1981) The Nereidae (Poluchaetous Annelids) of the Chinese Coast. Ocean Press, Beijing, 228 pp.

Yang D, Wu X, Wang Z, Zhao X, Hwang J, Cai L (2022) Redescription of a rarely encountered species *Travisachinensis* Grube, 1869 (Annelida, Travisiidae), including a description of a new species of Travisa from Amoy, China. ZooKeys 1128: 1–17. https://doi.org/10.3897/zookeys.1128.90020

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular ecology resources 20(1): 348–355. https://doi.org/10.1111/1755-0998.13096

Zhen W, Wu X, Hao E, Xu W, Deng J, Zhu J, Xu Y (2022) The first complete mitochondrial genome of the genus *Dendronereis*, represented by *D. chipolini* Hsueh, 2019 (Annelida, Nereididae) from Beibu Gulf, China. Mitochondrial DNA Part B 7(2): 393–395. https://doi.org/10.1080/23802359.2022.2040389

Supplementary material 1

The reports from FastQC

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Explanation note: (a) Basic information on clean data, including duplicate reads (%, Dups), average GC content (%, GC), and total sequences (millions, M Seqs). (b) Sequence counts for each sample. Duplicate read counts are an estimate only.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl1

Supplementary material 2

The report of Quast

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Explanation note: (a) Basic information of various assemblers. (b) The cumulative length of each assembler. All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl2

Supplementary material 3

Additional information

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: xlsx

Explanation note: **table S1.** The Blast results of mitochondrial gene, *18S*, *28S*, *ITS*, and histone genes (Only 20 sequences are shown). **table S2.** List of 32 species and two outgroups used in this paper. **table S3.** Sequence Information from NCBI BLAST Analysis of the *COX1* Gene. **table S4.** Sequence Information from NCBI BLAST Analysis of the *16S* Gene. **table S5.** Original and Gblock lengths of the PCGs and PCGsAA sequences. **table S6.** Best partitioning schemes and models based on different datasets for maximum likelihood and Bayesian inference analysis. **table S7.** *COI*, *16S*, *18S*, *28S*, *ITS*, and *H3* gene sequences information of *Perinereis*. **table S8.** Nucleotide composition and skewness comparison of different elements of the mitochondrial genomes of *P. wilsoni*. **table S9.** Features of the *P. wilsoni* mitogenome. **table S10.** Codon numbers and relative synonymous codon usage (RSCU) of 13 PCGs in the *P. wilsoni* mitogenome. **table S11.** Summary statistics for multiple alignment of various datasets.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl3

Supplementary material 4

The Python script for categorizing data

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl4

Supplementary material 5

Heterogeneity of sequence composition of mitochondrial genomes for 5 different data sets

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl5

Supplementary material 6

10 trees from different datasets and tree-building methods

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl6

Supplementary material 7

Phylogenetic trees of *Perinereis* based on the *COX1* dataset

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Explanation note: (a) Bayesian inference (BI, left) and (b) maximum likelihood (ML, right) method

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl7

Supplementary material 8

Phylogenetic trees of *Perinereis* based on the *16S* dataset

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Explanation note: (a) Bayesian inference (BI, left) and (b) maximum likelihood (ML, right) methods.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl8

Supplementary material 9

Phylogenetic trees of *Perinereis* based on the *ITS* dataset

Authors: Deyuan Yang, Sheng Zeng, Zhi Wang, Yanjie Zhang, Dazuo Yang, Christopher J. Glasby, Jiang-Shiou Hwang, Lizhe Cai

Data type: pdf

Explanation note: (a) maximum likelihood (ML, left) and (b) Bayesian inference (BI, right) methods.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.100.127201.suppl9